DOCUMENT RESUME

ED 068 355 24 SE 014 966

AUTHOR	Abrahams, Frederick F.
TITLE	Development of New Materials to Improve the Training of Students in Data Analysis. Final Report.
INSTITUTION	Brandeis Univ., Waltham, Mass.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
BUREAU NO	BR-8-A-051
PUB DATE	Feb 72
GRANT	OEG-1-9-080051-0001 (010)
NOTE	34p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Behavioral Science Research; *College Programs; *Course Organization; *Data Analysis; Educational Games; Evaluation Methods; Graduate Study; *Instruction; Methods Research; Research; Social Sciences; Statistical Surveys

ABSTRACT

New materials and approaches focusing on data analysis are described in this report. Students were exposed to a brief instruction on survey analysis. Then they used a series of games designed to aid the learning of data analysis, and worked with tables of data presented in different training formats designed to sharpen analytical skills. Limitations of previous methodology texts and instruction in data analysis are listed in the beginning of the report. Details are given of the training formats and the games which were developed, and their use is analyzed. A section on questionnaire construction is included. The evaluation of the effectiveness of the developed materials is given in terms of the author's impressions, based upon experiences with the materials in four graduate level methodology courses. (DT)

ED 068355

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

Final Report

Project No. 8-A-051
Grant No. OEC-1-9-080051-0001 (010)

Frederick F. Abrahams
Brandeis University
Waltham, Massachusetts 02154

DEVELOPMENT OF NEW MATERIALS TO IMPROVE THE TRAINING OF STUDENTS IN DATA ANALYSIS

February 1972

FILMED FROM BEST AVAILABLE COPY

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

National Center for Educational Research and Development

SE 014 966

ED 068355

Final Report

Project No. 8-A-051
Grant No. OEG-1-9-080051-0001 (010)

Development of New Materials to Improve the Training of Students in Data Analysis

Frederick F. Abrahams

Brandeis University

Waltham, Massachusetts 02154

February 1972

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
National Center for Educational Research and Development

TABLE OF CONTENTS

		Page
Chapter I	Background for the Study	1
	1. Introduction.	1
	2. Limitations of methodology texts.	3
	3. Limitations of instructional styles in methodology courses.	5
	4. Need for new textbook and classroom formats in methodology.	10
Chapter II	Methods	12
	5. Creating cross-tabulations.	12
	6. Formats for presenting tabular data.	13
	7. Games.	15
Chapter III	Findings	18
	8. Examples and analyses of games.	19
	9. Examples and analyses of formats.	22
	10. Questionnaire construction.	26
Chapter IV	Summary and Conclusions	27
	11. Summary.	27
	12. Conclusions.	31

I Background for the Study

1. Introduction

Today, the role of rationality in organization, in decision-making, and in the accumulation of knowledge is of paramount importance. Industrialization was the prime mover in elevating rationality to such exulted heights. Modern industrial societies constantly are attempting to apply rational criteria to more and more spheres of daily life. In our modern industrial societies, research has increasingly become the major intellectual method for reaching rational decisions. Accordingly, such societies devote ever greater amounts of time, personnel, and resources for research activities. Indeed, a hallmark of these societies is the support and faith they display towards research, both pure and applied. Nowhere is this devotion and encouragement of research more evident than it is here in the United States. The funds spent by the government, foundations, and private corporations on research in the United States annually run into the billions. Furthermore, the amount of time devoted to research activities by faculty, staff, and students at American universities is enormous.

Modern scientific research was initiated in the physical and biological sciences, and its application to agricultural problems gave rise to the development of modern statistics. In the last several decades, however, complex research designs have made their appearance in many other fields. In the United States a dramatic mushrooming of research activities in the social and behavioral sciences, as well as in such related fields as education, welfare, and administration, occurred during the post war era. Funds suddenly became available to research many aspects of social life, psychological phenomena, economic development, and political behavior.

In response to this unprecedented demand for research in these areas of the social and behavioral sciences, undergraduate and graduate curriculums have expanded their methodology offerings. In sociology, for example, undergraduate majors at the larger American universities are usually required to take at the minimum a one semester course in statistics and a one semester course in methodology. The minimal requirement for graduate students in sociology at many departments is two semesters each of statistics and methodology. In addition, such graduate students are often encouraged to enroll in one or more advanced courses focusing on survey methods, demographic analysis, field work or experimental design. In general, the methodology courses offered to both undergraduate and graduate students at American universities stress quantitative modes of analysis. In the fields of sociology and political science these methodology offerings focus on survey methods, while in psychology and education experimental methods predominate. In recent years, however, survey techniques have inroads in numerous additional fields, including educational research. Accordingly, an increasing number of methodology courses

offered at American universities devote at least part of their time to a discussion of survey research techniques.

Paralleling this expansion of methodology offerings in the social and behavioral sciences has been a corresponding expansion in the number of textbooks designed to be used in methodology courses. Most of these texts are geared to advanced undergraduates or first year graduate students and their scope is both broad and general. The topics discussed in such texts include the following: Philosophy of science, research design, procedures for data collection and processing, techniques preparatory to or concomitant with data analysis, data analysis itself, formats for the presentation of findings, and comments concerning the role of research. These texts cover survey and experimental methods, though in a brief and superficial manner. In addition to these general methodology textbooks, there are a few specialized texts intended for graduate level work that focus in depth on survey techniques.

The increasing exposure of social science students to methodology courses in which survey methods are covered and the rapid multiplication of texts available to students enrolled in such classes, imply that social science students are being adequately trained to carry out survey research. Unfortunately, this is not so. Both the procedures utilized in the classroom by instructors teaching survey research methods and the contents of most methodology texts that discuss survey techniques inadequately prepare students for the task of analyzing survey data. While some methods of research, such as experimental, clearly anticipate the detailed steps of analysis; other methods, such as survey, frequently fail during the design stage to foresee the analytical avenues through which the researcher will travel. Consequently, although a primary focus on the problems and techniques of research design is appropriate for training in experimental methods, it is insufficient when training students in the skills of survey analysis.

After the survey questionnaire has been designed, the interviews scheduled, responses recorded, and the data transferred onto IBM cards or computer tape, the real work of the survey researcher begins. He is often inundated with a large amount of data, and it is difficult to tell the "trees from the forest." Launching an analysis of survey data, which contain tens of thousands of tables or correlational coefficients, is like finding one's way out of an elaborate, confusing labyrinth. Many students never succeed in working their way out of such a maze. The survey analyst is faced with the task of finding a path through the labyrinth to reveal a coherent explanation of the phenomenon under investigation. The emphasis is thus on developing skills in <u>discovery</u> rather than on methods of <u>verification</u>. Unfortunately, students at American universities today are often confronted with classroom instruction and textbooks which focus only on methods of verification and fail to develop competence in their ability to <u>discover</u> relationships and to link

a series of relatively simple relationships into a complex, coherent structure. This project was an attempt to explore ways of filling this void.

2. Limitations of methodology texts

Despite the increase in the number and variety of methodology textbooks being turned out by American publishers, it is our contention that they fail to devote sufficient space to a discussion of how to proceed in analyzing survey data. This impression is borne out by a crude content analysis that we carried out of the material presented in approximately forty textbooks which were intended for use in first year methodology courses in sociology. Almost one-third of the pages of these textbooks were devoted to a discussion of the problems, procedures, and techniques which precede the collection of data -- e.g., developing hypothesis, research design, questionnaire construction, and sampling. This comprised the largest amount of space devoted to any one particular phase of the research enterprise. Another one-quarter of the pages of these textbooks was devoted to the problems, procedures, and techniques of analysis and presentation of the collected data -- e.g., general problems of measurement; construction of indices, scales, and factors; development of coding categories; statistical techniques; and the use of graphs, tables, and diagrams to summarize the presentation of research findings. While a mastery of these skills is a necessary prerequisite to becoming a competent survey analyst, in and of themselves, they are insufficient in aiding a student to launch his own analysis of survey data. They do not take him very far in his quest for a path out of the labyrinth of rich data with which he is confronted.

Approximately one-seventh of the pages of the texts surveyed focused on such questions as the relationship of research to theory, issues in the philosophy of science, and the implications and broader context of research. Another one-seventh of these pages was taken up with a discussion of how data is actually gathered -- e.g., interviewing and observational techniques. Finally, less than one-tenth of the pages in these texts was devoted directly to a discussion of how to analyze data. Furthermore, the presentation was rarely directed exclusively towards the most frequently used research method in American sociology -- namely, survey analysis. Even when the focus was on survey analysis, the discussion was frequently limited to an explanation of how to read two and three variable tables, rather than on how to discover relationships and how to link them together into a comprehensive analysis.

Up to this point, we have limited our discussion to general methodology textbooks designed for first year methods courses in

sociology. What about the more specialized methodology texts intended for graduate level courses and which exclusively deal with survey research? Do not such texts as Hyman's Survey Design and Analysis, Zeisel's Say It With Figures, and Rosenberg's The Logic of Survey Analysis adequately prepare the student for undertaking his own analysis of survey data? It is true that such texts do devote a considerable amount of space to a discussion of the actual steps involved in the analysis of survey data. Even these texts, however, emphasize the individual steps of data analysis in isolation to each other rather than in linking them together to form a comprehensive and smoothly flowing analysis. Thus, for example, these texts usually focus on how to read, analyze, and interpret individual two and three variable tables, as well as how to "elaborate" a simple two variable relationship by the systematic introduction of different third "test" variables. Little consideration, however, is given to developing the student's skill in linking a number of these tables together in order to form a comprehensive, relatively complex explanatory structure. This is a serious omission on the part of these texts since much survey research is initiated with the hope of coming up with such an explanatory structure for practical or theoretical purposes.

We will now shift our attention from the amount of space devoted to the actual analysis of data in methodology textbooks to a concern with the format through which instruction is given in these texts. One of three different formats, or modes of presentation, is employed by these texts. The first format emphasizes straightforward exposition of the methods and techniques covered in the text. Here, teaching is by talking. Goode and Hatt's Methods in Social Research exemplifies such a format. The second format emphasizes the exposure of the student to completed research reports representing well done research. Here, teaching is by example. Lazarsfeld and Rosenberg's The Language of Social Research illustrates this format. The third format is a hybrid combination of the two previous formats wherein presentations of abbreviated and edited research reports are followed by an explanation and discussion of the research techniques illustrated by these reports. Here, teaching is by example and talking. Riley's Sociological Research: A Case Approach is an example of this third format.

None of the three formats is too successful in aiding the student to develop skills at discovering relationships in data and in understanding how the survey analyst actually goes about his job, although the third is the least objectionable. The expositional presentations fail to get the student involved with actual research data while the presentations using completed research reports suffer from the fact that it is difficult, if not impossible, to reconstruct the actual steps taken by the researcher during his analysis from his formal presentation of the results. In other words, the format of reporting research differs considerably from the format used to carry it out.

In summary, while all of the topics covered in both general and specialized methodology textbooks should be mastered in order to carry out a research enterprise, the most crucial stage of that enterprise -- discovering relationships and linking them together into comprehensive and coherent causal chains -- is the one thing that these texts most neglect. Not only is relatively little space in these texts allocated for developing skills in data analysis, but, in addition, the formats utilized, in discussing data analysis, fail to enlighten the student about how research is actually carried out. Such texts thus emphasize methods of verification at the expense of methods of discovery, and they focus on the isolated building blocks of analysis without linking them together into a larger framework. The student, when confronted with real survey data, is often uncertain as to what to look for first, is undecided concerning what to look for next once he launches his analysis, and is unsure when to draw his analysis to a close. He is placed in the position analogous to a novice, at the game of chess, who has been instructed in the legitimate moves each chesspiece may make and what are the overall objectives of the game, but who lacks information about the appropriate strategy and tactics for opening game moves designed to develop his pieces, middle game moves to improve his position and engage his opponent, and end game moves designed to ensure victory.

Survey analysis is somewhat of a game, and the expert survey analyst, like the master chessplayer, must thoroughly understand the strategy and tactics appropriate to each stage of analysis and be able to anticipate several steps in advance the likely consequences of making one analytical move rather than another. Unfortunately, textbooks dealing with survey analysis, unlike chess manuals do not focus on building up the student's understanding of such strategy and tactics. Our project explored the feasibility of simulating actual research conditions to instruct students in research strategy and tactics, with the hope that these new approaches could be incorporated into future textbooks.

3. Limitations of instructional styles in methodology classes

Not only are textbooks negligent in presenting material that would aid students to develop the skills and insights necessary to carry out an adequate analysis of survey data, but classroom instruction techniques, also, fail in this regard. Upon completion of a methodology course that included instruction in survey research, students often found themselves still confused and ill at ease when confronted with actual survey data requiring analysis. They appeared unable to discover fruitful relationships which can be developed into more elaborate themes with interesting variations culminating in a coherent, comprehensive, and cogent explanatory system. The reasons for this failing in classroom instruction in survey analysis are several.

First and foremost among the reasons is the fact that little classroom time is given to instruction in the art of analysis itself. Just as methodology textbooks devote little space to a discussion of actual data, analysis instructors in methodology classes devote little time to the subject. Much more classroom time is usually spent in discussing general problems of research, research design, questionnaire construction, sampling procedures, interviewing techniques, coding operations and general problems of measurement, index construction, and scaling. Although all of these topics are related to the actual analysis of data, in and of themselves, they fail to give the student sufficient insight into the detailed steps of data analysis. More time should be allocated for instruction in the process of data analysis itself.

A second reason for the inadequate training of students is that what little time is spent on instruction in data analysis is at times abstract rather than concrete in nature and expositional rather than experiential. What the student sorely needs is practical experience in analyzing concrete survey data under the guidance of a trained instructor. Such practical experience is possibly less crucial in the teaching of experimental methods where the analysis is relatively systematic, codified, routinized, and anticipated in the research design stage. In survey research, however, where such attributes are regrettably lacking, practical experience is essential. Learning by doing is thus particularly relevant in the training of survey analysts. Unfortunately, in most methodology classes today, students spend very little time in doing survey analysis using concrete, real data. Moreover, even when such practice is undertaken by students, there is often a lack of adequate supervision by trained personnel.

A third and final reason why students in methodology courses are poorly prepared in the art of data analysis is that even when they are on occasion given practical experience in analyzing concrete survey data, that analysis is often confined to the simplest level. Thus, students are encouraged to look at the univariate distributions of the responses given to the various questions by the respondents and then proceed to detect relationships between some of these variables by constructing two variable tables. At times, students go somewhat further in "elaborating" these two variable building blocks of survey analysis by the introduction of three variable tables. Often the entire analysis is limited to a handful of two variable tables supplemented by one, two, or three variable tables. Even at this elementary level students are often confused and not given sufficient time and supervision to develop their competence and confidence in data analysis. Moreover, they often gain no insight whatsoever into the more demanding task of linking a number of component two and three variable tables together in order to form a comprehensive and coherent analytical structure of sufficient scope to tell an interesting sociological "story." The strategies and tactics required for such a task are rarely communicated to the

student. Thus, to resort to our analogy with chess-playing, the student is at best limited to an understanding and practice of the "opening" moves of the analysis, and he is relatively ignorant of the appropriate moves for the further, "middle" development of his inquiry and of the "end" moves that will successfully wind up the analysis.

In our earlier discussion of textbooks, we indicated the different formats used to convey the material and the limitations of each of these formats. Similarly, we will now discuss the different formats used by instructors in methodology courses and indicate the limitations of each. In general, there are four different formats of classroom instruction in methodology courses that we have identified. One relies solely on lectures and discussion; another supplements such lectures and discussions with simple practice exercises; a third calls for the involvement of the students in a major class project which attempts to be a microcosm of a professional survey operation from beginning to end; and a fourth has students engaging in the secondary analysis of previously collected survey data. We will now discuss each of these four formats in more detail.

The first format we will consider calls for the teaching of methodology courses along the lines traditionally used for most other college courses. Here the instructor lectures on a single topic each class meeting. Thus, for example, one or several class meetings are devoted exclusively to problems of research design, or on questionnaire construction, or on measurement techniques. These lectures are often supplemented by discussion sessions led by the instructor or an assistant during which the students can raise questions concerning points covered in the lectures that perplex them. Hopefully, these discussion sessions serve to resolve any confusion in the student's mind as well as supply him with supplementary information on the topics covered by the lectures.

This approach to the classroom teaching of methodology suffers in several respects. To begin with, since each topic is considered in isolation of the others, it is difficult for the student to appreciate the integrated nature of the research enterprise. Instead, he often emerges with a segmented and disconnected image of research activity, failing to see how decisions at one stage have repercussions on all the subsequent stages. In addition, this format fails to give the student any practice in analyzing survey data. Students in such courses are often at a loss when asked to carry out even a simple analysis of survey data summarized in tables. They are often bored or frustrated. Such feelings might be partially overcome if students are introduced to the more exciting, more creative aspects of research which involve the discovery of interesting relationships and linking them to form a causal explanation of some interesting phenomena.

Format number two is an elaboration of the first. Here, the instructor supplements the lectures and discussion sessions with a series of practice exercises designed to give the student more direct contact with different phases of the research. The students, for example, might construct simple questionnaires, interview one another, code open-ended responses to a previously administered questionnaire, draw a mock sample, see if a set of items form a scale, or analyze a few selected tables. Although these activities on the part of the students are not integrated or utilized in an actual research project, they, nevertheless, give the students some tangible contact with specific research operations.

This format, also, exhibits several serious weaknesses. Like the first format, this also presents the research enterprise as a series of isolated operations whose integration is not experienced by the students. While students taught according to this second format do engage in practice exercises, these exercises are often too brief, artificial, and segmented for students to develop much insight or skill in research. Furthermore, too little time is spent on practice exercises in data analysis, and the students are almost as much at a loss to carry out even simple analyses of survey data as those students taught in courses using format number one. Finally, many of the practice exercises are felt to be boring or irritating and are carried out in a reluctant and resentful mood. This further detracts from whatever heuristic value the practice exercises may have.

The third format entails having students collectively carry out a simplified version of a complete research project. Students, together with the instructor, frame a research problem that is susceptible to investigation with survey techniques, develop a research design that translates the problem into appropriate research steps, construct an interview schedule, draw a sample to be interviewed, administer the schedule to respondents, code the responses to the interview, punch the coded information onto IBM cards, arrange for the transfer of the data onto computer tape, write programs instructing the computer as to how to process the data, analyze the computer print-outs, and write a report based on the analysis. Here, the student does, indeed, acquire practical experience in carrying out a survey. He gains experience with individual research operations, and, more important, he begins to see how they are integrated.

Although this format of instruction is superior to the two preceding formats, it, also, has its limitations. Class projects such as these are limited to the larger colleges and universities with equipment and funds. This approach is impractical for use by the large number of small colleges. Rather than use computers for data processing, they must rely on hand sorts on McBee cards, or utilize punched card equipment. Furthermore, little time is spent in the actual analysis of the data by students engaged in such

collective projects. The greatest bulk of their time is spent on constructing a questionnaire, interviewing, preparing the data for analysis, and learning how to run punched card equipment and/or writing computer programs. As a result, the final reports turned in by students from such projects are often of low quality reflecting the haste and confusion with which they were put together. Finally, the brevity of the interview schedule and the small number of respondents that usually characterize such class projects often prevent the students from emerging with a penetrating analysis of the phenomena under investigation.

The fourth and final format calls for students to undertake an extensive secondary analysis of previously collected survey data. The survey data used is usually of professional calibre -- the research design, the interview schedule devised, the drawing of the sample and interviewing of the respondents, and the processes by which the data are prepared for processing. Professional standards in these matters means that there are fewer "errors" at each stage of the research process. The cumulative effect of many errors that could seriously undermine the credibility of the analysis is avoided. Such a danger is more likely to occur in a class research project. Since the data used for a secondary analysis is often a large scale survey, the range of information solicited from the respondents and the large number of respondents interviewed permit much greater sophistication in the analysis. Secondary analysis focuses exclusively on the analysis phase of research and encourages the student to discover new relationships in the data and to link them in order to form an integrated, comprehensive, and smoothly flowing analysis. In conclusion, this format gives the student the greatest exposure to the experience of analyzing real survey data. Therefore, it is the most useful of all the formats in developing skills of data analysis.

Despite the obvious advantages of this format, it, too, has its drawbacks. To begin with, it fails to emphasize the integration of the various steps of the research enterprise. The student may fail to appreciate all the effort and skills that ultimately led to the creation of the data that he is using. Although more time is spent by the student in actually analyzing data in courses using this format than in those using any of the previously described formats, a substantial amount of time is still occupied with non-analytical tasks which tend to distract one's attention from analyzing tables -- e.g., learning machine operation and programming, understanding the codebook, and construction of analysis decks or tapes. Still another drawback is the fact that even more so than in the case of class projects, secondary analysis is mainly limited to the larger, more affluent colleges and universities where modern data processing equipment is available for student use. It, also, depends on the possession by the instructor of the raw data from a previous study that is in satisfactory condition to be analyzed by students. Moreover, personnel are required to assist the students

in punched card machine operation and/or computer programming. Few small colleges are likely to have the equipment, material, and personnel needed to carry out secondary analysis. Finally, the data used in secondary analysis would have been previously analyzed and, in all likelihood, published by a professional researcher, and the most obvious lines of inquiry would have already been explored. The student would be forced to focus his attention on questions that are peripheral to the original research design. The task becomes more demanding than would be the case if students were given an unanalyzed study. The difficult task all novices face in coming to grips with survey data for the first time is further compounded by the need to restrict their inquiry to the more hidden, secondary paths.

4. Need for new textbook and classroom formats in methodology

The foregoing sketch of the limitations and failings of both classroom instruction and textbook presentation of methodology highlights the need for new approaches. The most cruicial need today in the training of students in methodology is for new approaches that help the students develop greater understanding and insight into the actual process of data analysis. Such new approaches are particularly needed in the training of students in survey research where the research design rarely completely anticipates the subsequent lines of analysis. This, we believe, can best be accomplished by having the student attempt to discover relationships in real or simulated data and then theoretically link such discovered relationships together into a comprehensive and coherent explanation of the phenomena under investigation. These new approaches, then, should focus predominantly on the art of data analysis itself rather than on auxiliary stages of the research process. Furthermore, they should focus on the logic of discovery rather than on the logic of verification. In addition, these new approaches should be capable of holding the students' interest and stimulating their curiosity. Finally, they should be able to be employed in small as well as in large colleges and universities. In other words, these new approaches should be designed to compensate for the inadequacies existing in the approaches currently employed in methodology texts and course.

The development of competence in the analytical skills and insights that characterize the professional survey analyst can be fostered, according to some, by extensive participation in ongoing research projects. According to this view, the only form of training that is at all useful is an apprenticeship type program wherein the student adopts the role of a research assistant. The apprenticeship model of "learning by doing" has much to be said for it, especially, when the activity of the student is closely supervised by a professional researcher. Nevertheless, the major assumption upon which our project rested was that alternatives to the apprenticeship model could be developed and incorporated, both in methodology texts and classroom presentations. These alternatives should approximate, to some extent, the "learning by doing" experience associated with the

apprenticeship model while at the same time having wider application. They should supplement the approaches already employed in methodology courses. These new approaches should aim at reducing the wide chasm that currently exists between the training students receive in survey analysis and the analytical skills and insights actually needed to carry out a competent analysis of survey data. A reduction in this chasm would have two important consequences. It would encourage a larger number of students to participate in survey research projects in the future, and it would enhance the value these students would derive from such participation.

The first consequence -- that of encouraging greater participation in survey research -- would come about because students would be introduced at an early point into the most exciting and creative aspect of the research enterprise -- data analysis. Presumably, this would result in greater enthusiasm and excitement about research on the part of students which, in turn, would encourage them to become apprenticed to professional researchers carrying out ongoing survey research projects. The second consequence -- that of deriving greater value from participating in future survey research -- would result from the fact that students would more likely engage in the analysis of data, rather than in being relegated to routine and less challenging tasks as interviewing, coding, or percentaging.

The recent development of survey data banks, both here and abroad, calls for an increasing number of persons equipped to engage in detailed secondary analysis of data bank material. The approaches developed in this project are designed to help the student become competent in secondary analysis.

II Methods

5. Creating cross-tabulations

The basic assumption underlying the approaches developed in this project is that students best learn data analysis by actually doing it. While discussions of the general principles underlying data analysis and exposure to examples of completed research reports are of value, they are no substitute for the actual analysis of data itself. Often, however, it is impractical or inconvenient to plunge novices immediately into the confusion of the swelling sea of raw interview data. Consequently, we have developed comprehensive sets of "middle-range" data which fall between the extremes of the raw data contained in interview schedules (which are transferred onto IBM cards or computer tape) and the highly selected, polished, statistical or tabular presentations found in published research reports.

The "middle-range" data consists of a large number of conventional tables, each representing the cross tabulation of two or three variables. These tables were intended to approximate the range of tables an investigator might well have ordered in the process of analyzing raw survey data. Some of these tables should prove to be of little value in the subsequent analysis, while others should be of crucial value. This, again, approximates the situation sometimes encountered by professional survey analysts whose initial exploratory lines of analysis prove, upon investigation, to be useless. The talented survey analyst can pull together those few tables that convey the thrust of his analysis, from the enormous number of tables that potentially can be ordered. Our "middle-range" data can aid the student develop his ability to tell the "trees from the forest."

We developed four different sets of such "middle-range" tabular data. Each set focused on a different substantive area. These four areas were (1) the voting behavior, (2) race relations, (3) student attitudes and behavior, and (4) social mobility. We felt that this was preferable to developing tables dealing in only one area, as it increased the likelihood that students would find an area of personal interest, which would elevate the students' motivation and involvement. These particular four areas were selected because of our familiarity with each and also because we possessed raw data from previous surveys in each of the four areas.

The tables for each area were created in three ways. (1) Some were derived by cross-tabulating variables contained in real surveys; (2) others were purely inventions of our imagination, based on no real survey data but guided by our knowledge of what certain relationships would look like; and (3) still other tables were constructed by starting with real data and then "fudging" the results slightly in order to highlight connections for heuristic purposes. Each table, regardless of how it was generated, contained the frequency distributions, both in terms of the number of cases falling into each category and the percentage distribution computed

columnwise. In addition, each table contained several statistical measures of association.

6. Formats for presenting tabular data

We developed two different formats within which the "middle-range" sets of tables were presented to the students. In all cases our ultimate goal was to have the students formulate some research problem related to the data they would receive. They were, then, to proceed to discover relationships between variables cross-tabulated in tables and link selected tables to form an integrated analysis and interpretation. In this section, we will describe how we created those two formats. In the following section, we will discuss the relative success and failures of these two formats.

Format number one began by presenting the student with a final research report of professional calibre. Some such reports were created especially for this project, while others were taken from the professional literature. Reports gathered from the latter source were either presented as originally published or were altered somewhat for heuristic purposes. All of these reports, regardless of their source, contained numerous tables interspersed within the text. The report was immediately followed by some commentary concerning alternative possible interpretations of the tables. An attempt was made to indicate why some such interpretations were preferable to others. This, in turn, was followed by a mythical discussion between myself and the author of the report, and the presumed path taken in the analysis was recapitulated step by step. The discussion helps the student appreciate the many false starts, dead ends, changes in direction, and un-anticipated turn of events that take place in real-life research. The student begins to understand that the orderly and logical sequence of findings presented in the final report is a poor reflection of the mental processes and data manipulation that take place during the ongoing research enterprise.

We, then, elaborated on different ways in which the original data could have been integrated, and alternate interpretations of the phenomena under investigation resulted. Some selected tables were juxtaposed for the purpose of illustrating some of these alternative lines of inquiry. Thus, the student was presented with concrete examples of alternative versions of the final report.

Next, the student was supplied with a set of two and three variable tables, which were derived from the same data that were the basis for the tables appearing in the original report. He was instructed to do two things with these tables. First, he was to select the ones which best supported the alternative lines of inquiry that we suggested were possible. Here, the student is given a "story" and has to fit some data to it in the best possible fashion. Second, given the set of tables, he was instructed to develop some alternative lines of inquiry of his own. Here, the student is given some data and has to uncover a "story" contained therein. The student, together

with the instructor, then, compares and contrasts the various alternative "stories" developed, and criteria are presented for deciding which of these alternatives are superior to others.

The students were then supplied with an extensive set of tables, either real or fictitious, which dealt with a topic related to the one discussed in the original report. There were, then, several exercises that the students had to perform with this data. To begin with, limited hypotheses and rudimentary "scripts" were supplied along with the set of tables, and the students were told to discover which tables, if any, supported these hypotheses and "scripts." The students were also requested to develop their own hypotheses and "script" outlines after looking at a few tables. In both cases, the students were then instructed to look at additional tables that would either tend to confirm or deny the initial hypotheses and "scripts." At first, they were told to hold on to these hypotheses tenaciously. They were to explain away as best as they could any empirical data that contradicted these hypotheses, as well as to search energetically for additional empirical confirmation. Next, they were told to discard gradually some of these hypotheses and formulate new ones if the empirical data appearing in the tables failed to substantiate them. Finally, they were told to discard these hypotheses the very first time some empirical data failed to support them and formulate alternative hypotheses. These exercises were intended to convey to the student the capricious nature of the link between theory and data.

The last step was to have the student use the set of tables given him for the purpose of developing an extensive "story." The student was to link a number of empirically substantiated hypotheses into a coherent interpretation of the phenomena under investigation. He was, also, told to list any additional tables which he would like to have included in order to further substantiate his analysis.

The approach taken in this first format was to guide the student in gradual stages from the finished report back to raw data in tabular form. We started with a well structured and organized interpretation and by stages introduced greater entropy until the student was brought face to face with disconnected tables of data. The second format we employed proceeded in the opposite fashion. We initially exposed the student to a large set of raw data in tabular form and then in gradual stages guided him to the finished report created by a professional researcher. The student begins with an extensive set of tables, scans through them, and starts to develop a "story" outline suggested by the data. He, then, is requested to develop more limited hypotheses and "scripts" from this data. He is, then, given several feasible hypotheses and "scripts" which could be derived from the data which he then compares with those he previously created. The student is then required to put together a sequence of tables that will best support particular

hypotheses and "scripts" that are supplied him. Next, the student is confronted with a reported reenactment of the mental processes a professional researcher might go through in trying to tease a "story" out of the set of tables supplied. Finally, the actual finished report is read by the student. In brief, format two is really format one stood on its head.

Each format contained some novel approaches in the teaching of data analysis. We were concerned both with how these different aspects went over as well as with the question of which of the two formats were more effective teaching devices. In the following section on findings, we will attempt to answer these question. Before we do this, however, let us turn to several other heuristic devices we developed.

7. Games

In addition to the formats discussed above, we also developed several heuristic "games" designed to introduce students to the art of data analysis. We purposely inserted a competitive aspect in most of these games in order to heighten motivation and interest. The students played at the various games before being exposed to the two more extensive training formats. Prior to playing these games, all students received rudimentary instruction in survey analysis. They were instructed in the following concepts and techniques: Distributions, cross-tabulations, correlations, "elaboration," and multivariate analysis. Instruction at this stage was more condensed and focused than is usually the case, and we purposely ignored a number of issues normally discussed. Our intent was not to have the students thoroughly master these various aspects of data analysis, but merely to gain a superficial familiarity with the terms and logic of survey analysis. A more thorough understanding of these matters, we hoped, would come about after exposure to the material we had developed.

Following this brief instruction period, the class was broken up into small groups, and each group was divided into two teams. These teams then played several games we had developed expressly for heightening their insight into the art of survey analysis. A script containing a "story" based on data and empirical interpretations had been given to each team for study before class convened. One team was instructed to try its best, by use of the empirical interpretation tables (which were on cards), to substantiate the interpretive story line developed in the script; the other team was instructed to try its best to refute that story line.

During the class period, both teams were to discuss the matter, and each team would attempt to win over the other. Each group of two teams made its presentations to the entire class. One team argued for the script and the data; the other attempted to demonstrate that the data was ill suited for supporting the story line emerging from the script. After both teams had engaged in thorough debate, the

instructor led the class in a discussion on the merits of the arguments presented by each side.

The winning team received a certain number of positive points; while the losers were penalized with negative points. If, prior to the class presentation, one team had wisely altered its position in light of the other team's arguments, it was not penalized. On the other hand, if the class as a whole felt that its original position had merit and was defendable, the team that had shifted from that position was penalized to a greater extent than it would have had it remained steadfast in its convictions. After the class had reached a judgment concerning the cases presented by each team from the first group, two teams from a second group made their presentation. This routine continued until all of the groups were heard. This process often took several consecutive class meetings. The winners of one group played the winners of another group; while the losers of each of these two groups played against each other. This continued until most teams had accumulated a certain set number of positive points in the scoring. Once a team had reached a predetermined scoring level, indicating its competence in elementary analysis and presentations, its members were free to move on to the more elaborate variations of the game. There were several such variations.

Other games we developed did not require team competition. One such variation required each team to enunciate alternative "scripts" that might have been written to fit the tables. They were then judged according to how believable their script was compared to the original. A second variation called for each team to receive at regular intervals, additional tables, and they were to determine whether or not these additional tables altered the fit between the "script" and the data, and, if so, in what ways. A third variation gave each team several scripts and a large number of tables. Each member of a team was given a script, which he read aloud; while the other members searched through the tables in an effort to find those that would substantiate the story line contained in the scripts.

Another was our "domino" game. In this game, the class was again divided into small groups. Each of these groups was given a set of related cards of one particular sociological phenomenon and which, also, represented the relationship between different pairs of variables. Each student in the group then received an equal number of cards, and the game began.

The game commenced by placing a special octagonally shaped card that contained four pairs of two-variable relationships in the center of the table. The first student attempted to "match" one of the cards he held in his hand to one of the eight sides of the center card. Then, each student in turn would attempt to add one of his cards to the emerging structure created by the linking of sequence cards. One end of each card represented one of the two variables in the relationship; while the other end represented the second variable.

All of the cards were limited to representing two variable relationships. The center of the card indicated the nature of the relationship. This relationship was indicated in two different ways: (1) by correlational coefficients and (2) by cross-tabulations. One could add to an already discarded card only by joining a variable at one of its ends with an identical variable in card one still held. If during his turn, a player had no card and he felt he could add to the emerging structure, he would then pick up one or more additional cards from a common pool of cards until he could discard one. Thus, the play proceeded along lines similar to those used by children in playing dominoes. Unlike dominoes, however, it was possible to break apart two previously connected cards for the purpose of inserting a new card between them. The positioning of the new card in relation to these cards already discarded indicated different implications in the flow of causality. The "winner" of the game was the first person who discarded all the cards he had held.

The students were to verbalize about the "story" they were spelling out in the domino game. For example, one card may show a relationship between religious affiliation and voting behavior. A second card, attached to the first, indicates that there is a relationship between class and partisanship, while a third attached card shows that there, also, is a relationship between class and voting behavior. The students might interpret this to mean that class factors intervene in interpreting the original relationship between religion and partisanship. At this point, however, it would be pointed out to them that they would need a three-variable table to substantiate this interpretation.

A separate pool of three-variable tables was made available to each group from which the students could then search for the appropriate table that would support their interpretation.

Our last game involved the creation of three different sets of cards. One set of cards contained two and three variable tables; a second set was composed of analyses of these tables; and a third set was made up of more general interpretations of these analyses. The class was divided into several small groups of three students. In each group, one student received the table cards; a second, the analysis cards; and a third, the interpretation cards. The student with the table cards would begin by discarding one of his cards. Next, the student with the analysis cards would search for the one which best analyzes the table and proceeds to discard it. Finally, the third student would attempt to discard one of his cards which adequately interprets the analyzed table. After all the cards were played, the students would exchange sets of cards and begin again.

III Findings

How effective were the various devices we developed in heightening the students' comprehension and skill in the art of data analysis? In order to answer this question, we initially intended to enlist the aid of colleagues at different universities who were teaching courses in survey methods courses in either sociology or political science departments. They were to use the materials in a selective way whenever possible, so that a clearer picture of their effectiveness might emerge. For example, students in a class were to be divided into a number of different sub-groups on a random basis, and different materials were to be used by the students in each sub-group. All the students in a class were to be given identical examinations designed to test their ability to analyze actual sets of tables. In addition, the students were to engage in a brief analysis of conventional raw survey data. The relative abilities of the different students in data analysis, as measured by both their grades on the standardized tests and their performances in analyzing actual survey data, were to be correlated with the different materials to which they were exposed. Thus, our evaluation of the relative effectiveness of the materials which we had developed depended to a large degree upon the systematic feedback of information from colleagues utilizing these materials in methodology classes.

We should point out that even with our own classes, limitations of time and personnel made it impossible to rigorously follow the "testing" format suggested above. Hence, the findings reported in this section largely rest on our subjective impressions concerning the relative effectiveness of the materials presented. Nevertheless, considering the "exploratory" nature of this project in exposing students to a variety of novel approaches in the teaching of data analysis, our impressionistic findings may not be a liability.

Flexibility, rather than rigidity, is important at the exploratory stages of a project. This flexibility facilitates the discarding of approaches which seemed promising earlier but fail to live up to expectations. Flexibility also encourages tinkering with the original material in light of experiences with it in actual classroom use, so as to improve its value. It, also, promotes the spontaneous development of new, unanticipated materials suggested by our classroom experiences. A preconceived experimental testing design mitigates against such flexibility in favor of scientific rigor. Such rigor is more appropriate at a later stage when we want to choose from among several alternative sets of materials, which seem from impressionistic exploratory probes to have some value.

Let us now turn to the impressions we formed concerning the effectiveness of these materials in training students in the art of data analysis. We will begin by focusing on our experiences with the

several data analysis games we developed and discuss the effectiveness of the two training formats.

8. Examples and analyses of games

In all our methodology classes, the students were briefed on the rudiments of survey analysis before proceeding to our specially designed games. How well did this instruction go over? The first trial class reeled from having too much information thrown at them. The students felt that they had not coped adequately with one topic before they were in the midst of discussing another. With our second class, we reduced the amount of material introduced in the beginning. The topics that had been eliminated were introduced later at intervals. Furthermore, rather than begin by discussing methodology points in the abstract, we began by posing concrete problems in different substantive areas of sociology. We then proceeded to show the students how different types of data handled in different ways can throw light upon these substantive problems. This problem-oriented approach proved much more effective than the abstract approach in introducing students to the essentials of survey analysis. At this stage, the students had not mastered the art of survey analysis. Nevertheless, they had acquired an appropriate vocabularly, understood the rudimentary logic of survey analysis, and began to appreciate the value of methodological points in clarifying real world problems, both practical and theoretical.

Approximately six meetings of our methodology classes were taken up with a discussion of these basics in survey analysis. The students were then introduced to the games. How well did these games go over? What alterations were made in light of our experiencws? And how effective were these games in realizing our purpose? The games, by and large, were well received. Students appeared to enjoy the competitive and novel aspects of the games. With our novice students, these games were more successful in holding interest than the conventional methods of teaching data analysis. Let us now turn to a discussion of what alterations we made in these games.

The first game involved dividing the class into several groups of two teams each. Each group received a script, which elaborated a story line, and a set of tables, which formed the empirical basis of the script. The teams played against each other. One team tried to demonstrate that the data presented in the tables supported the story line in the script, and the other team attempted to show that the story line was unsubstantiated by the data. What impressions did we form on the basis of our experiences with this game?

First, this game consumed much more time than we had intended. In subsequent trials, we reduced the amount of time needed by introducing the following alterations: (1) The groups were made larger, so that there were fewer teams making their presentations in class;

(2) the discussion between the teams in each group now took place outside of the regular classroom, either in laboratory sessions or at someone's residence; and (3) we divided the class into two parts, and teams made their presentations simultaneously to one or the other half of the class, rather than have every team present its arguments before the entire class. As a result, the total class time consumed by this game was kept within reasonable bounds.

A second impression of our experiences with this game in the early trial stages was that the students' class presentations were weak and confusing. We felt this was due partially to insufficient time allowed for discussing and preparing the presentation and to a lack of supervision in developing such a presentation. In order to increase the time each team spent outside of the class discussing the script and the tables, we had more copies reproduced. This enabled every team to have its own copy of both the script and the tables, and meetings were arranged more readily and frequently than had been possible with the larger groups. In order to give the students supervision in preparing their presentations, each team met at least twice with a teaching assistant. The assistant participated more as a group facilitator and information source rather than as an authoritative leader. These changes brought about higher calibre presentations in succeeding classes.

Third, the involved point scoring system which we had developed proved confusing and irritating to the students and time consuming for us. After a while, we had the teams play against one another without keeping a record of their scores. Nevertheless, "winners" of each contest played against one another in a round-robin fashion until one team had triumphed over all previous "victors." The members of that team then went on to play some of the other games among themselves, while the other teams continued to contest each other until each was victorious over the remaining teams. The competitive spirit was retained, but the burdensome point system was discarded. This method allowed the "losing" teams to have more practice in playing the game while the "winning" teams enjoyed the opportunity to play the more individualized games.

Finally, we noted that some of the teams were dissatisfied with the assignment to either defend or refute the story line contained in their script. During later trials, we gave each team the option of deciding among themselves which position they would like to defend after they had had an opportunity to inspect both the script and the tables. Although this procedure resulted in greater interest and motivation, it suffered from two drawbacks. In the first place, an unequal number of teams chose a particular position; consequently, the pairing of teams was difficult to arrange. Secondly, too much time was consumed by teams deciding which of the two positions they wished to defend.

In the future, a compromise arrangement would be best. Each team would be assigned a position, then allowed a brief time in which to decide whether or not they would chose to defend that position. If not, they could alter their position if they would find another team to trade position with them. In this way, it should be possible to maximize interest and still have an equal number of opposing teams.

After a team had defeated the remaining teams in the round-robin contest, its members began to play the other individualized games among themselves. The first of these called for the students to create new scripts from the tables which their team had received earlier. Each student was to do this independent of the others, and after each had composed such scripts, he would discuss the strengths and weaknesses of the scripts with the other members of the group. Despite several attempted variations, the students experienced considerable difficulty in creating appropriate scripts. Thus, at this stage in their training, students were much more hesitant about spinning stories from data than they were in relating data to stories which had already been constructed.

All teams were later given additional tables, and each member of the team had to indicate in which ways the additional data might alter the previous fit between tables and script. Some of these additional tables were purposely designed to refute the story line developed in the script; others were designed to buttress that story line. This exercise sharpened the analytical eye of the student and made him appreciate the way in which additional information can alter one's empirically grounded notions. Although the students had some difficulty with this exercise at first, they later did very well and appeared to enjoy it thoroughly.

The next game involved having each member of a team read out a script, and the other members searched through a pool of tables to substantiate the story line developed in the script. This game worked reasonably well. We added the requirement that the students also search for tables that would refute the story line contained in the script. After some trials, we arrived at the conclusion that the best way to utilize this game was to divide the members of each team into three groups. One team was comprised of those who would attempt to refute the story line; a second team, of those who would try to substantiate it; and a third composed of that individual who was reading the script. In order to facilitate the processing of the tables, we gave each team two sets of the tables, one for each group arguing for or against the story line developed in the script.

The students were later divided into small groups, and they played our domino game using as dominoes, cards embodying two variable relationships pertaining to a single sociological phenomenon. This game caused some perblems among the students at first. We made a number of alterations. To begin with, rather than have a whole group play together, we restricted the play to two students at a time. This

caused less confusion among the students, reduced the waiting time between successive moves by the same student, and increased the likelihood that tables that one student needed were not already appropriated by others. Second, we had all the cards in the pool turned face up, so that students could pick only those cards that they wanted. This reduced the number of cards a student held in his hand at any one time. This, in turn, reduced the student's confusion. Third, the hands dealt to each player were pre-determined by us, rather than left to chance. This made it easier for each player to develop a story line with the cards in hand. Finally, to each pair of teams playing the games, we assigned one of the better players. His function was to pass judgment on disputes as well as to be in charge of the set of three variable tables. We are still not completely satisfied with the current state of this game, but its usefulness appears much improved from what it was at first.

The last game involved the use of cards symbolizing the three different levels in the processing data -- tables, analysis of tables, interpretation of the analysis. One student would discard a table-card; the next student would throw out an analysis-card that corresponded to that table-card; and the third student would discard a card which best interpreted the analysis of the table. The students performed quite well at this game. Nevertheless, they appeared bored and as if they felt they had mastered these skills earlier. We believe that this game should be the first, rather than the last, one encountered by the students.

How effective were all of these games after above alterations and modifications had been made? In our judgment, they were very effective. After completing the entire series of games, most students appeared to understand the ways in which survey data can be used to confirm or disprove hunches, hypotheses, and theories. They also appeared knowledgeable in the ways in which a sequence of tables can form the empirical foundations for an involved explanation of a sociological phenomenon. Furthermore, they gave every indication of being at ease with tables and data, rather than intimidated and awed -- a characteristic reaction among novice methodology students. When several selected students at this stage "ordered" tables from actual data that was computer processed, their understanding, competence, and confidence were considerably above those of students exposed only to conventional survey analysis training for the same period of time. However, the students still had difficulty starting with a set of tables and developing a story line from them. This skill, we hoped, would become developed when they were exposed to our training formats.

9. Examples and analyses of formats

The students were given ample opportunity to play the variable games. Afterwards, they were exposed to the two formats we had developed earlier. The purpose of these formats was to sharpen

the analytical skills that students developed in the process of playing the games. These formats showed how middle-range survey data, i.e., tables, were integrated in order to form a finished research report. The roles of assumptions, analytical skills, and interpretive flare were emphasized. The students were shown how tables are woven in gradual stages. One format began with the report and ended with the data; while the other format began with the data and culminated in the report. After a student read through the format and engaged in some of the exercises, he pursued his own analysis of the large set of supplied tables. Let us now discuss our impressions based on our experiences in developing and using these training formats.

The tables, as well as the reports, were created in one of three ways -- out of our imagination, out of actual data and reports, and combination of the two. In our judgment, the third alternative (combination of imagination and actual data and reports) worked out best in classroom use. When we limited ourselves to using actual data and reports, it was difficult to find exactly what we wanted in order to make methodological points. Relationships between variables were rarely striking enough to catch the eye of the novice analyst, and it took considerable amount of time and effort to grind out the appropriate tables that would serve our purposes from the raw survey data. Nevertheless, the students appeared enthusiastic about the idea of using real data. They felt they were learning about real people. They were less enthusiastic when fictitious data and reports were used. Creating these fictitious tables and reports proved time consuming, and, at times, we created tables which were not consistent despite efforts to avoid this pitfall. Heuristically, however, these tables and reports were ideal for making methodology points clear to the students. The third alternative -- that of starting with actual tables and reports and the altering them slightly -- combined the advantages of the first two alternatives. Student enthusiasm remained high, the time spent in preparing the materials was less than in the other two procedures, and for the most part, the tables were consistent with one another, and the methodology points we wanted to make were clear.

We propose that in the future our middle range tabular data be prepared in only one, rather than in four, substantive areas. The reasons are several. To begin with, it took considerably more time to develop these formats than we had anticipated. When faced with the task of developing four such formats, our time and effort were divided, and not enough attention was paid to developing a format. If we limit ourselves to creating only a single format, less total time should be absorbed in this task, and more attention and care could be devoted to a thorough construction of this single format. Also, interaction among the students, instructors, and teaching assistants should be enhanced by having the entire class work on only one format. Teaching assistants could be used more effectively in dealing with questions when the entire class is working on a single format.

A number of the students complained that, among the set of tables we supplied, there were too many useless ones. Few useless tables should be included among the set of tables the student is to analyze at first. This will minimize his confusion and of being overwhelmed by the data. Once the student gains some confidence in his ability to analyze the data, he can be given another set of tables which does contain many useless tables. In light of our experiences with these formats, we, also, feel that at the early stages, students should work with only a few tables. The number of tables can be gradually enlarged as his confidence and skill grow.

The first time we utilized these formats, we asked the students to pay attention both to the percentage point differences and the correlational coefficients contained in each table. This proved confusing to the students and hindered rather than aided the development of their analytical skills. Subsequently, we had the students focus exclusively upon the percentage point differences contained in each table ignoring all other statistical measures. Only after the students were able to demonstrate that they could competently analyze the tables and spin off empirically grounded interpretations were they told to pay attention to the correlational coefficients. They were then told to retrace their analytical steps, relying solely on these statistical measures. This enabled them to comprehend the relative advantages and disadvantages of correlational coefficients compared to percentage point differences. Finally, they were to analyze the tables by looking at the correlational coefficients alone. They were then to repeat their analysis having recourse to the percentage point differences shown for each table. These exercises were very useful in helping the students develop an appreciation of the strengths and weaknesses of different analytical procedures. These exercises, however, consumed a considerable amount of time.

The students looked with favor upon the discussion style of presentation between the supposed author of the report and ourselves during that stage when we attempted to recapitulate the actual creative process that culminated in the final report. They were somewhat overwhelmed and confused by the extent and variety of the analytical "gig-gags" that took place in the research process. This was, also, the case when we confronted them with numerous alternative lines of investigation that could have been followed by the author if he had chosen. It is our feeling that this confusion can be reduced substantially in the following ways: (1) Spelling out fewer alternative lines of investigation or analytical turns and twists by the author of the report and (2) interspersing the discussion between the presumed author of the reports and ourselves with pertinent tables that would empirically illustrate and highlight the alternative turns that were taken.

There appeared to be no clear cut preference on the part of the students for one type of format over the other. Some students who were given formats that led them from data to the final report expressed regret that they were not assigned the other format. On the

other hand, a like number who received formats that started with the final report and ended with data similarly thought that they would prefer the alternative format. Most students appeared content with the type of format they received. From our perspective, it did not seem that one type of format was more effective than another in developing the students' analytical skills.

How effective were these formats in realizing their dual objectives -- of (1) increasing the student's awareness and understanding of the intellectual turns and twists that link the final report of an actual research enterprise with the raw data initially collected and (2) improving the student's skill in interpreting complex, causal models emerging out of the analysis of tables? In our view, the formats were very successful in achieving the first objective and somewhat less successful in realizing the second. The students did, indeed, grasp the fact that the order of the findings presented in a final report bore but a faint resemblance to the intellectual gyrations and data manipulations that occurred in the research process. They began to appreciate the fact that research involves a great deal more than the mere following of set routine procedures by which findings are ground. They, also, started to understand that the art of research is much more creative, challenging, and absorbing, as well as frustrating, than they had originally imagined. Finally, the students began to see the importance of the researcher's implicit assumptions and explicit decisions on the outcome of the final report. The role of such assumptions and decisions were not clearly appreciated by the students during their initial reading of the final research report.

The formats, also, improved the student's own ability to link data with interpretations. Many students still encountered difficulties, however, when they attempted to spin off elaborate and coherent stories on the basis of their data analysis. However, their performance in this regard was clearly superior to that of students who received only traditional instruction. The formats were effective in improving the student's ability to move from raw data to polished report, but not quite as effective as we had hoped.

Performance may improve if after working through the formats, each student is given a carefully selected set containing only a few tables. Then, after he demonstrates his competence in developing interpretations of different combinations of these tables, he would be given a few additional tables to add to his set. In gradual stages, he should be encouraged to develop more comprehensive and involved interpretations and should not be permitted to advance to the next stage until he has demonstrated his ability to carry out a satisfactory analysis with the tables at his disposal. This procedure should build the student's confidence and encourage him into handling larger and larger files of data.

10. Questionnaire construction

The games and formats discussed above were designed to aid the student in developing analytical understanding, insights, and abilities. This was, after all, the major objective of our project. Initially, we had not planned to develop any materials designed to improve other phases of the research -- e.g., interviewing, questionnaire construction, or coding. It became clear to us that it would be important to sensitize the students to the impact the construction of questionnaires have on the analysis of the collected data. We developed several exercises for this purpose.

First, students were given a set of tables and asked to substantiate given story lines where either (1) important pieces of information were absent because they failed to have been asked or (2) it was impossible to carry out crucial tests concerning alternative explanations because vital questions had not been asked during the interview. In these cases, the process of analysis serves to point out additional new questions that should have been asked of the respondents. The students are told to look out for such negligence in their analysis and to write up such questions and insert them in the printed questionnaire that is supplied them. Students were, also, encouraged to make up additional questions that should have been asked in the interview that would have improved their analysis. In another exercise, students were told about a problem area and asked to design a brief questionnaire. They were asked to write out a request for the tables they would want in order to launch an analysis based on these questions. The teaching assistants then made up some mock tables that gave empirical flesh to their skeletal request form. The students were to analyze the tables and then indicate what additional questions should have been inserted into their questionnaire. This exercise shows the difficulty encountered in anticipating the questions to ask at the questionnaire design stage. The students were also given the results of slightly differently worded, but similar questions from different surveys taken at different dates. The difficulties encountered there in arriving at reliable trend analysis illustrated the results obtained with slight alterations in wording.

These exercises proved effective in sensitizing the student in the relationship between data analysis and questionnaire construction. Awareness of the problems in questionnaire construction was heightened. The students no longer saw these problems as isolated. They began to appreciate the advantages of anticipating some of the analysis during the questionnaire design stage.

IV Summary and Conclusions

11. Summary

A hallmark of modern society is the support and faith it displays towards research. During the past three decades the United States witnessed a dramatic growth of research activities in the social and behavioral sciences as well as in such related fields as education, welfare, and administration. In response to this unprecedented demand for research in these fields, the curriculums geared to social science majors at both undergraduate and graduate levels have noticeable expanded their methodology offerings, and there has been a corresponding increase in the number of methodology textbooks designed for social science students. The increased exposure of social science students to methodology courses coupled with the rapid multiplication of methodology textbooks might lead one to conclude that the social science students are being adequately trained in the art of data analysis. Unfortunately, this presumption is unwarranted. This is particularly true in the case of survey analysis, which today constitutes one of the major forms of social research.

Unlike experimental data, the precise path through which the survey analyst travels is not adequately foreseen in the research design stage. The survey analyst finds himself inundated with data from which tens of thousands of potentially useful tables can be created. Telling the "trees from the forest" is an essential talent in such a situation. The survey analyst seeks a path through the elaborate and confusing maze of data that will lead to a coherent and revealing description or explanation of the phenomenon under investigation. The search for such a path calls for the development of insight into the methods of discovery rather than to methods of verification. Unfortunately, classroom instruction and textbook presentation in American universities today often focus exclusively on methods of verification. They fail, by and large, to develop the student's insight into how to go about discovering relationships and linking them to form complex and coherent explanatory structures. The development of such insights, however, is an essential ingredient in the training of novices in the art of survey analysis. New materials and approaches are needed if such training is to be effective. This project is a first attempt at creating such new materials and approaches for the training of students in the art of survey analysis. The project is exploratory rather than definitive in nature.

Today's student of survey data is often at a loss as to how to proceed. He is uncertain about how to launch his analysis; he is undecided on how to develop his analysis once he has initiated it; and he is unsure about when it is appropriate to draw his analysis to a close. He is like a novice chessplayer who merely receives some elementary instruction concerning the legitimate moves allowed each chesspiece along with a brief introduction concerning the overall

objectives of the game. The game played by such a chess novice is poor, because (1) he lacks information concerning strategy and tactics -- i.e., opening game moves designed to develop his pieces, middle game moves designed to improve his position and engage his opponent, and end game moves designed to ensure victory -- and (2) he has no understanding of how to integrate these stages. Like the chess master, the professional survey analyst must thoroughly understand the strategy and tactics for each stage of analysis, anticipating several steps in advance.

New materials and approaches are needed to develop the student's insight into the strategy and tactics appropriate to the different stages of the research enterprise. This project created such novel materials and approaches and then proceeded to try them out on social science students. These materials and approaches were designed to compensate for the inadequacies present in current American methodology texts and courses. These new materials and approaches have the following characteristics. They focus on the art of data analysis itself rather than on auxiliary processes in the research enterprise. In the area of data analysis, they concentrate on the art of discovery rather than on the logic of verification. They give the student ample opportunity to engage in the actual analysis of data rather than on being expositional in nature. They hold the student's interest and involvement, and they are suitable for use in small as well as in large colleges and universities.

The new materials and approaches we developed were based on the assumption that the student best learns data analysis by engaging it it. Discussion of the principles or exposure to examples of research are no substitutes for the analysis of data itself. It is often, however, impractical and inconvenient to plunge novices into the confusion of the sea of raw interview data. As an alternative, we provided the student with sets of two and three variable tables as their basic data input. These tables fall between the extremes of raw data contained in interview schedules and polished final research reports.

These tables were created in three different ways. Some were the products of our imagination; some were taken intact from actual survey data; and some were initially based on survey data, but altered for heuristic purposes. Tables created in this last manner proved to be the best. Different sets of tables were created in the following four substantive areas -- voting behavior, race-relations, student attitudes and behavior, and social mobility. As a result of our experiences in using these materials in methodology courses, we are of the opinion that in the future it would be wiser to restrict the use of such material to a single substantive area. These tables were extensively used in our training games and formats that will be discussed. These games and formats constitute the novel materials and approaches developed by us.

We had hoped to enlist the voluntary help of colleagues in evaluating the effectiveness of the materials we had developed. We distributed copies of our materials, along with evaluating tests, to a number of colleagues in sociology or political science departments at different institutions where they were teaching survey research methods courses. Unfortunately, we have not received the feedback from these colleagues. As a result, this report is largely limited to our own impressions based on our experiences with the materials in four graduate level methodology courses. This approach, although inappropriate for rigorous testing, has value in an exploratory project, such as ours. It allows for flexibility which facilitates the discarding of approaches that fail to fulfill promise. It stimulates experimentation with the prepared materials, and it promotes spontaneous development of new materials suggested by those experiences.

All students were initially exposed to a brief instruction on the rudiments of survey analysis. The content of this instruction was pared down to the bare essentials, and the format of presentation was changed from an abstract one to a problem-oriented approach. The students did not master the art of survey analysis in the six (around six) meetings devoted to this basic instruction. Nevertheless, they developed a preliminary understanding of the logic of survey analysis along with an appreciation of its value in illuminating problems existing in the real world. In addition, they acquired a basic vocabulary that facilitated their reading of methodology articles.

After being introduced to the ABC's of survey analysis, the students were exposed to a series of games, which were developed by us to initiate students into the art of data analysis. They utilized the sets of tables described earlier, as well as scripts containing story lines pertaining to these tables. In the first game, the class was divided into several groups, and each group was split into two teams. Each group was given a set of tables and a script, and the two teams of each group played against each other. One team tried to demonstrate that the data presented in the tables supported the story line developed in the script; while the other team attempted to show that the story line was unsubstantiated by the data. The two teams gave presentations before the entire class, which judged on which team made the better presentation. This game was well received by the students.

Our experiences suggested several ways in which the amount of game time could be reduced. By increasing the size of each group, we reduced the time needed for classroom presentations. By insisting that discussion among members of each team, as well as debates between teams, take place outside of the classroom, valuable class time was saved. Finally, presentations were made only to half, rather than the entire class; therefore, two separate presentations could occur simultaneously. The elaborate point scoring scheme designed to foster competition between teams was replaced by a round-robin format. This format retained the competitive element but eliminated the confusion

emerging from the involved point-scoring system. This game could be improved by giving each team an option on whether or not it wishes to defend the tables.

After a team had emerged "victorious" from the round-robin contest, its members played other data analysis games among themselves. One such game was similar to the game discussed above -- i.e., matching sets of tables to scripts -- except that in this case, competition took place among members of a single team rather than between teams. Another game involved giving each team additional tables at intervals and having them indicate ways in which the additional tables altered the previous fit between data and script. Still another game called upon students to create scripts based on their analysis of the tables. The students found it more difficult to spin stories from data than to search through a set of tables in order to substantiate a prepared story. Another game sought to sharpen the students' awareness of the relationship between tables, analyses, and interpretations was one in which one student received table-cards and proceeded to discard them one at a time. A second student, supplied with analysis-cards, tried to match each table-card with an analysis-card. A third student, given only interpretation-cards, attempted to find an appropriate interpretation-card. It is our judgment that this game would be more effective if it were used earlier, rather than late, in the training period. The final game was our domino game. Students received cards representing the relationship between different pairs of variables. The complete set of cards related to a single sociological phenomenon. Although this game was initially designed for a group of players, we decided to restrict play to only two players at a time. One player would throw out a card; the other player tried to match one of his cards to the discarded cards. Eventually, a structure of cards would be formed, like a pattern of dominoes, with the players giving verbal interpretations of the connected tables.

After all the students had had opportunity to play these games, they were exposed to training formats. These formats were designed to further sharpen the analytical skills, and they demonstrated to the students how tables were woven together into a final research report. This demonstration proceeded step by step through several stages -- (1) the polished research report, (2) an unveiling of the actual analytical path pursued by the author of the report, (3) a discussion of alternative paths that could have been pursued by the author, (4) an opportunity for the student to discover for himself some other alternative paths of analysis, and, finally, (5) giving the student a comprehensive set of tables to analyze and weave into a report. One format began with the final report and traced the stages backward, ending with the raw data; while a second format began with the data and culminated in the report. In practice, neither format was deemed superior to the other. Some formats were created from actual research reports; some, from our imagination; and some began as actual reports but were altered for heuristic purposes. Formats created in this latter fashion worked out best in classroom use.

The games and formats were designed to familiarize the student with the strategy and tactics of survey analysis. We initially chose to ignore the other phases involved in the research enterprise. In practice, however, we decided to sensitize the students to the effect of questionnaire construction on the subsequent analysis of data. In order to do this, several exercises were developed. Students were asked to spin off stories from sets of tables derived from questionnaires which lacked crucial questions. The students were required to write up additional questions that would improve the analysis. In another exercise, the students created a brief questionnaire of their own. Mock tables were created by us on the basis of their questionnaire to highlight the absence of certain questions. Finally, the students were exposed to the different results obtained when slight rewording of questions takes place. In total, these exercises made the students more aware of the problems of questionnaire construction and its relationship to data analysis.

12. Conclusions

By and large, the students were enthusiastic about the games. Our flexible approach stemming from the exploratory nature of this project encouraged us to make numerous changes in these games for improving their use. These games, as modified, were indeed effective in introducing students to the art of survey analysis. Students seemed to understand the ways in which survey data could be utilized for the purposes of confirming or rejecting hunches, hypotheses, and theories. They also appeared to fathom the ways in which combinations of tables can form the foundation of complex explanations and interpretations of sociological phenomena. Rather than being intimidated by the tables, most of the students appeared at ease with the data. Finally, the analysis that several students carried out on actual computer processed survey data was of a much higher calibre than that carried out by students exposed only to conventional training in survey analysis. Finding tables to corroborate a given story line presented no problems; however, the students still had difficulty developing a story line from tables. It had been hoped that the new formats which we developed would have remedied this weakness.